JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 17, 161-171 (1967)

On a Multi-Stage Nonlinear Programming Problem

S. ARIMOTO

Department of Mathematical Engineering, The Faculty of Engineering, University of Tokyo, Tokyo, Japan

Submitted by Richard Bellman

1. INTRODUCTION

In optimization problems of multi-stage processes or discrete-time control processes, some types of necessary conditions for optimality were proposed in several papers, like the maximum principle for continuous-time control processes. In the earliest paper [1], Chang proposed a necessary condition for optimality and called it "the digitized maximum principle." Similar results were also obtained by Katz [2]. Butkovsky [3], in addition to a counter-example to Katz's theorem, offered the local maximum principle which implies that the Hamiltonian attains the maximum value in a neighborhood of the optimal control. However, the local maximum principle for discrete-time processes does not hold in general as shown by a counter-example which will be given in the last section of this paper.

Recently, the two important papers by Halkin [4], Jordan and Polak [5] were published. Halkin showed geometric aspects of necessary conditions for optimality, and Jordan and Polak established the *local maximum or stationary* principle.

In the present paper we shall consider more general optimization problems for multi-stage processes than those in the above cited references.

The problem stated in Section 2 is a discrete version of Berkovitz's problem [6] formulated for continuous-time control systems. It is also regarded as a generalization of nonlinear bottleneck-type programming problems in multi-stage production processes first discussed by Bellman [7]. In Section 4 a necessary condition for optimality will be proved.

Sections 5 and 6 treat a special case of the problem. In Section 5 a sufficient condition for local optimality will be given in Theorem 2. In Section 6, a global maximum principle will be proposed in Theorem 3 under an additional condition which is analogous to that given by Fillipov [8] for the proof of the existence of an optimal control for continuous-time systems.

In our previous work [9] we proposed the analogous theorem to Theorem 3

409/17/1-11

given in Section 6, but, in the proof, we falsely used the local maximum principle which does not hold in general.

2. Problem Statement

Let us consider a multi-stage process whose state at the tth stage ($t = 0, 1, \ldots$) or time t is described by an n-vector x_t governed by the difference equation

$$x_{t+1} = f_t(x_t, u_t), \tag{2.1}$$

where u_t is an r-vector called a decision and f_t is an n-vector valued function which has continuous first derivatives with respect to all arguments of x_t and u_t. Given an initial state x_0 and a sequence $u = \{u_t;\, t = 0, 1, \ldots, N-1\}$ of decisions, there exists a unique solution of (2.1) denoted by $x_t = x_t(x_0, u)$.

The problem to be considered is

Problem 1. Given an initial state x_0, find a sequence of decisions $u_0, u_1, \ldots, u_{N-1}$ which minimizes

$$J(N; x_0, u) = \sum_{t=0}^{N-1} \alpha_t(x_t(x_0, u), u_t), \tag{2.2}$$

subject to the process Eq. (2.1) and the inequality side constraints

$$g_t^{(i)}(x_t(x_0, u), u_t) \geqslant 0, \qquad i = 1, 2, \ldots, m \tag{2.3}$$

for $t = 0, 1, \ldots, N-1$, where $g_t^{(i)}$ has continuous first derivatives with respect to all arguments.

Throughout this paper we assume that there exist at least one sequence of decisions and the corresponding sequence of states $x_t(x_0, u)$ along which (2.1) and (2.3) are satisfied. We shall call such a sequence of decisions to be admissible.

3. Preliminary Formulations

Let v_t be an m-vector and define

$$\begin{aligned} \beta_{t+1} &= \beta_t + \alpha_t(x_t, u_t) + v_t' g_t(x_t, u_t), \\ \beta_0 &= 0, \end{aligned} \tag{3.1}$$

where the prime denotes the transpose and $g_t(x_t, u_t)$ is the m-vector whose components are composed of $g_t^{(i)}(x_t, u_t)$ in (2.3). Let $u^* = \{u_t^*\}$ be a fixed admissible sequence of decisions and denote the corresponding state by $x_t^* = x_t(x_0, u^*)$ and the solution of (3.1) by $\beta_t^* = \beta_t(x_0, u^*)$. Then, after a lengthy but easy calculation, we obtain

$$x_{t+1} - x_{t+1}^* = \frac{\partial f_t(x_t^*, u_t^*)}{\partial x_t^*}(x_t - x_t^*) + \frac{\partial f_t(x_t^*, u_t^*)}{\partial u_t^*}(u_t - u_t^*) + h_t(x_t - x_t^*) + k_t(u_t - u_t^*) \tag{3.2}$$

and

$$\begin{aligned}\beta_{t+1} - \beta_{t+1}^* - (\beta_t - \beta_t^*) = {} & \frac{\partial[\alpha_t(x_t^*, u_t^*) + v_t' g_t(x_t^*, u_t^*)]}{\partial x_t^*}(x_t - x_t^*) \\ & + \frac{\partial[\alpha_t(x_t^*, u_t^*) + v_t' g_t(x_t^*, u_t^*)]}{\partial u_t^*}(u_t - u_t^*) \\ & + h_t^0(x_t - x_t^*) + k_t^0(u_t - u_t^*),\end{aligned} \tag{3.3}$$

where h_t, k_t are n-vector valued functions and h_t^0, k_t^0 are scalar functions such that

$$\begin{aligned} | h_t(x_t - x_t^*) | &= 0(| x_t - x_t^* |), \\ | k_t(u_t - u_t^*) | &= 0(| u_t - u_t^* |), \\ | h_t^0(x_t - x_t^*) | &= 0(| x_t - x_t^* |), \\ | k_t^0(u_t - u_t^*) | &= 0(| u_t - u_t^* |). \end{aligned} \tag{3.4}$$

Here the symbol $| x |$ implies the norm of the vector x, i. e.,

$$| x | = \max_j | x^{(j)} | ,$$

and the symbol $| h(x)| = 0(| x |)$ implies that for an arbitrarily given $\epsilon > 0$ there exists a positive number ρ such that $| h(x)| \leqslant \epsilon | x |$ for every x satisfying $| x | \leqslant \rho$.

Next we introduce the following notations:

$$H_t(x_t^*, p_t^*, u_t) = - \alpha_t(x_t^*, u_t) + p_t^{*\prime} f_t(x_t^*, u_t), \tag{3.5}$$

$$F_t(x_t^*, p_t^*, u_t) = H_t - v_t' g_t(x_t^*, u_t), \tag{3.6}$$

where p_t^* is an n-vector determined by

$$p_{t-1}^{*\prime} = \frac{\partial F_t(x_t^*, p_t^*, u_t^*)}{\partial x_t^*}, \tag{3.7}$$

together with the boundary condition

$$p_{N-1}^* = 0. \tag{3.8}$$

Finally, we define

$$d_t(u, u^*) = \beta_t - \beta_t^* - p_{t-1}^{*\prime}(x_t - x_t^*). \tag{3.9}$$

Then it follows from (3.2) and (3.3) that

$$d_{t+1}(u, u^*) - d_t(u, u^*) = - \frac{\partial F_t(x_t^*, p_t^*, u_t^*)}{\partial u_t^*}(u_t - u_t^*)$$
$$+ h_t^0 + k_t^0 - p_t^{*\prime}(h_t + k_t). \tag{3.10}$$

Note that

$$\begin{cases} d_0(u, u^*) = 0, \\ d_N(u, u^*) = \beta_N - \beta_N^*. \end{cases} \tag{3.11}$$

4. Necessary Condition

Let

$$U_t(x) = \{u; g_t^{(i)}(x, u) \geqslant 0, i = 1, \ldots, m\}$$

and define

Condition 1. For an arbitrarily fixed x, $U_t(x)$ is a convex subset of R^r.

Condition 2. If $g_t^{(j)}(x, u) = 0$ for $j = i_1, \ldots, i_k$ and for arbitrarily fixed x, u, then it holds

$$\operatorname{rank}\left(\frac{\partial g_t^{(j)}(x, u)}{\partial u^{(i)}}\right) = k.$$

Theorem 1. *Assume Conditions* 1 *and* 2. *Let* $u^* = \{u_t^*\}$ *be an optimal admissible sequence of decisions which minimizes* (2.2). *Then for every* $t = 0, 1, \ldots, N-1$ *it holds that* (i) *there exists an m-vector* v_t *which satisfies*

$$v_t^{(i)} g_t^{(i)}(x_t^*, u_t^*) = 0 \qquad i = 1, \ldots, m, \tag{4.1}$$

$$\frac{\partial F_t(x_t^*, p_t^*, u_t^*)}{\partial u_t^*} = 0, \tag{4.2}$$

(ii)
$$v_t^{(i)} \leqslant 0 \qquad i = 1, \ldots, m, \tag{4.3}$$

(iii)
$$\frac{\partial H_t(x_t^*, p_t^*, u_t^*)}{\partial u_t^*}(u_t - u_t^*) \leqslant 0 \tag{4.4}$$

for all $u_t \in U_t(x_t^*)$.

PROOF. At first we assume that the part (i) holds for every t. Assume that for $t = s + 1, \ldots, N - 1$ the conclusions (ii) and (iii) hold but for $t = s$ do not. Then, by Lemma 1 stated later, there exists a decision $\bar{u}_s \in U_s(x_s^*)$ such that

$$\frac{\partial H_s(x_s^*, p_s^*, u_s^*)}{\partial u_s^*} (\bar{u}_s - u_s^*) = \rho > 0. \tag{4.5}$$

Let λ be a small positive number and

$$u_s(\lambda) = \lambda \bar{u}_s + (1 - \lambda) u_s^*.$$

Then we have $u_s(\lambda) \in U_s(x_s^*)$ by Condition 1 and

$$\frac{\partial H_s(x_s^*, p_s^*, u_s^*)}{\partial u_s^*} (u_s(\lambda) - u_s^*) = \lambda\rho > 0$$

by using (4.5). Consider now the following process:

$$\begin{aligned} x_{t+1}(\lambda) &= f_t(x_t(\lambda), u_t(\lambda)) \qquad t = s, s+1, \ldots, N-1, \\ x_s(\lambda) &= x_s^*. \end{aligned} \tag{4.6}$$

Here, $u_t(\lambda)$ is determined successively such that

$$v_t' g_t(x_t(\lambda), u_t(\lambda)) = 0 \qquad t = s + 1, \ldots, N - 1. \tag{4.7}$$

We note that, from Lemma 2 stated later, it is possible to choose $u_t(\lambda)$ such that, in addition to (4.7),

$$u_t(\lambda) \in U_t(x_t(\lambda)), \qquad | u_t(\lambda) - u_t^* | \leqslant c\lambda \tag{4.8}$$

for $t = s + 1, \ldots, N - 1$, where c is a positive constant independent of λ. Thus, if λ is sufficiently small, we have from (3.1), (3.8)–(3.11) that

$$\begin{aligned} J(N - s; x_s^*, u^*) &= \sum_{t=s}^{N-1} \alpha_t(x_t^*, u_t^*) \\ &= \sum_{t=s}^{N-1} [\alpha_t(x_t^*, u_t^*) + v_t' g_t(x_t^*, u_t^*)] \\ &= \beta_N(\lambda) - d_N(u(\lambda), u^*) \\ &= \sum_{t=s}^{N-1} [\alpha_t(x_t(\lambda), u_t(\lambda)) + v_t g_t(x_t(\lambda), u_t(\lambda))] + 0(\lambda) \\ &= J(N - s; x_s^*, u(\lambda)) + v_s' g_s(x_s^*, u_s(\lambda)) + 0(\lambda). \end{aligned}$$

Noting that

$$v_s' g_s(x_s^*, u_s(\lambda)) = v_s' g_s(x_s^*, u_s^*) + \frac{\partial v_s' g_s(x_s^*, u_s^*)}{\partial u_s^*} (u_s(\lambda) - u_s^*) + 0(\lambda)$$

$$= \frac{\partial H_s(x_s^*, u_s^*)}{\partial u_s^*} (u_s(\lambda) - u_s^*) + 0(\lambda)$$

$$= \lambda\rho + 0(\lambda),$$

we get

$$J(N - s; x_s^*, u(\lambda)) = J(N - s; x_s^*, u^*) - \lambda\rho + 0(\lambda)$$
$$< J(N - s; x_s^*, u^*)$$

by choosing λ small enough. This contradicts the optimality of the sequence u^*. Finally, the proof of the part (i) follows from Lemma 3 stated later.

Now, it remains only to prove the following lemmata.

LEMMA 1. *If the inequality* (4.4) *holds for all* $u_t \in U_t(x_t^*)$, *then* (4.3) *holds.*

PROOF. For simplicity we omit the subscript and asterisk. We assume, without loss of generality, that for $j = 1,..., m_0$, $g^{(j)}(x, u) = 0$ and for $j = m_0 + 1,..., m$, $g^{(j)}(x, u) > 0$. Note that by Condition 2 there exist the vectors $u^1, u^2,..., u^{m_0}$ such that

$$\frac{\partial g^{(j)}(x, u)}{\partial u} (u^k - u) = \begin{cases} 1 & \text{for} \quad j = k, \\ 0 & \text{for} \quad j \neq k, \end{cases} \qquad k, j = 1, 2,..., m_0 .$$

To prove this lemma by contradiction we assume that $v^{(k)} > 0$ for some k. Let

$$u^k(\lambda) = \lambda u^k + (1 - \lambda) u,$$

$$\bar{u}^k(\lambda) = (1 - \gamma) u^k(\lambda) + \delta \sum_{j \neq k} u^j(\lambda),$$

where γ and δ are small positive numbers such that $\gamma = (m_0 - 1)\delta$. If λ is taken small enough, then

$$\frac{\partial g^{(j)}(x, u)}{\partial u} (\bar{u}^k(\lambda) - u) = \begin{cases} (1 - \gamma) \lambda & \text{for} \quad j = k, \\ \delta\lambda & \text{for} \quad j \neq k. \end{cases}$$

This implies $\bar{u}^k(\lambda) \in U(x)$. Hence

$$\lambda(1 - \gamma) v^{(k)} + \lambda\delta \sum_{j \neq k} v^{(j)} = \frac{\partial H(x, u)}{\partial u} (\bar{u}^k(\lambda) - u) \leqslant 0$$

by using (4.2). The last inequality follows from the assumption of the lemma. On the other hand, the left-hand side of the above equation becomes positive by choosing δ sufficiently small. Thus the contradiction has been derived.

LEMMA 2. *Assume that*

$$v^{(j)}g^{(j)}(x, u) = 0 \qquad \text{for} \qquad j = 1,..., m \tag{4.9}$$

and

$$| x(\lambda) - x | \leqslant M_0\lambda, \qquad u \in U(x) \tag{4.10}$$

for all λ such that $0 \leqslant \lambda \leqslant \rho$, where M_0 and ρ are positive constants. Then for any sufficiently small λ there is a decision $u(\lambda)$ such that

$$v'g(x(\lambda), u(\lambda)) = 0 \tag{4.11}$$

and

$$| u(\lambda) - u | \leqslant M_1\lambda, \qquad u(\lambda) \in U(x(\lambda)), \tag{4.12}$$

where M_1 is a positive constant independent of λ.

PROOF. Without loss of generality, we assume that for $j = m_0 + 1,..., m$, $g^{(j)}(x, u) > 0$, and for $j = 1,..., m_0$,

$$g^{(j)}(x, u) = 0. \tag{4.13}$$

Keeping in mind of Condition 2 and the property (4.10), and applying the theory of implicit functions to equation (4.13), we find that there exists a $u(\lambda)$ such that

$$g^{(j)}(x(\lambda), u(\lambda)) = 0 \qquad j = 1,..., m_0$$

and

$$| u(\lambda) - u | \leqslant M_1\lambda.$$

On the other hand, if λ is sufficiently small, we have

$$g^{(j)}(x(\lambda), u(\lambda)) > 0 \qquad j = m_0 + 1,..., m.$$

These imply (4.11) and (4.12).

LEMMA 3. *Assume that for given x and u,*

$$g^{(j)}(x, u) = 0 \qquad j = 1,..., k,$$

where $k \leqslant m$, and

$$F(x, p, u, v) = H - \sum_{j=1}^{k} v^{(j)}g^{(j)}(x, u).$$

Then there exist $v^{(j)} (j = 1, ..., k)$ *such that*

$$\frac{\partial F(x, p, u, v)}{\partial u} = 0,$$

PROOF. The proof of this lemma follows immediately from Condition 2.

5. SUFFICIENT CONDITION

In this section and the subsequent we consider the special case of Problem 1 when $U_t(x)$ is independent of x, namely, x_t does not enter the constraint inequality (2.3). Hence we say that a sequence $u = \{u_t\}$ of decisions is admissible if every u_t belongs to the set U_t which is a subset of R^r.

COROLLARY. *Assume that* U_t *is a convex set for all* t. *Let* $u^* = \{u_t^*\}$ *be an optimal admissible sequence of decisions minimizing* (2.2). *Then it holds for all* $t = 0, 1, ..., N - 1$ *that*

$$\frac{\partial H_t(x_t^*, p_t^*, u_t^*)}{\partial u_t^*} (u_t - u_t^*) \leqslant 0$$

for all $u_t \in U_t$.

Now we introduce the following notions.

DEFINITION. If for an admissible sequence $u^* = \{u_t^*\}$ of decisions there is a positive number ϵ such that it holds

$$J(N; x_0, u^*) \leqslant J(N; x_0, u) \tag{5.1}$$

for all admissible sequences $u = \{u_t\}$ satisfying $| u_t - u_t^* | \leqslant \epsilon$ for $t = 0, 1, ..., N - 1$, we say that the sequence u^* is *locally optimal.*

DEFINITION. In addition to the local optimality, if the equality symbol in (5.1) occurs only when $u = u^*$, we say that u^* is *locally strict-optimal.*

THEOREM 2. *Let* $u^* = \{u_t^*\}$ *be an admissible sequence of decisions and assume that for all* $t = 0, 1, ..., N - 1$ *it holds*

$$\frac{\partial H_t(x_t^*, p_t^*, u_t^*)}{\partial u_t^*} (u_t - u_t^*) < 0 \tag{5.2}$$

for any u_t such that $u_t \in U_t$, $u_t \neq u_t^*$ and $| u_t - u_t^* | \leqslant \rho$, where ρ is some positive constant. Then the sequence u^* is locally strict-optimal.

PROOF. We prove this theorem by contradiction. Assume that there are infinite admissible sequences of decisions $u(k) = \{u_t(k)\}$, $k = 1, 2, \ldots$, such that

$$0 < | u(k) - u^* | = \max_t | u_t(k) - u^* | \leqslant \frac{1}{k} \tag{5.3}$$

and

$$J(N; x_0, u(k)) \leqslant J(N; x_0, u^*), \tag{5.4}$$

and denote the corresponding states by $x_t(k) = x_t(x_0, u(k))$. Let

$$\frac{\partial H_t(x_t^*, p_t^*, u_t^*)}{\partial u_t^*} (u_t(k) - u_t^*) = - c_t(k) < 0.$$

At first we note that it follows from the assumptions (5.2) and (5.3) that

$$| u_t(k) - u_t^* | \leqslant M_1 c_t(k) \leqslant M_2/k,$$

where M_1 and M_2 are positive constants independent of t and k. Hence, by the same proceeding as in the proof of Theorem 1, we obtain

$$J(N; x_0, u^*) = J(N; x_0, u(k)) - \sum_{t=0}^{N-1} c_t(k) + 0\left(\sum_{t=0}^{N-1} c_t(k)\right).$$

Thus we have

$$J(N; x_0, u(k)) > J(N; x_0, u^*)$$

if k is sufficiently large. This contradicts (5.4).

REMARK. It should be noted that the conclusions in Theorem 1 and Corollary are valid even if u^* is locally optimal.

6. GLOBAL MAXIMUM PRINCIPLE

We requires the following property.

CONDITION 3. For all x, any given $u^1, u^2 \in U_t$, and any positive number $0 \leqslant \lambda \leqslant 1$, there exists at least a decision $u^3 \in U_t$ such that

$$\begin{aligned} \lambda f_t(x, u^1) + (1 - \lambda) f_t(x, u^2) &= f_t(x, u^3), \\ \lambda \alpha_t(x, u^1) + (1 - \lambda)\, \alpha_t(x, u^2) &\geqslant \alpha_t(x, u^3) \end{aligned} \tag{6.1}$$

for every $t = 0, 1, \ldots, N - 1$.

Now we prove the global maximum principle.

THEOREM 3. *Assume Condition 3. Let $u^* = \{u_t^*\}$ be an optimal admissible sequence of decisions. Then it holds for all $t = 0, 1, \ldots, N-1$ that*

$$H_t(x_t^*, p_t^*, u_t^*) \geqslant H_t(x_t^*, p_t^*, u_t) \tag{6.2}$$

for all $u_t \in U_t$.

PROOF. Let $u = \{u_t\}$ be an arbitrarily fixed admissible sequence of decisions and $\lambda = \{\lambda_t\}$ be a sequence such that $0 \leqslant \lambda_t \leqslant 1$ for $t = 0, 1, \ldots, N-1$. Let

$$x_{t+1} = \lambda_t f_t(x_t, u_t) + (1 - \lambda_t) f_t(x_t, u_t^*) \tag{6.3}$$

and denote the state vector for given x_0 and $\lambda = \{\lambda_t\}$ by $x_t(\lambda) = x_t(x_0, \lambda)$. We now consider the new optimization problem of choosing an optimal sequence $\lambda = \lambda^* = \{\lambda_t^*\}$ such that it minimizes

$$J(N; x_0, \lambda) = \sum_{t=0}^{N-1} \alpha_t(x_t(\lambda), \lambda_t). \tag{6.4}$$

subject to $0 \leqslant \lambda_t \leqslant 1$. Of course, it follows immediately from the meaning of this problem that

$$\min_\lambda J(N; x_0, \lambda) \leqslant J(N; x_0, u^*). \tag{6.5}$$

On the other hand, we have from Condition 3 that

$$\min_\lambda J(N; x_0, \lambda) = J(N; x_0, \lambda^*) \geqslant J(N; x_0, u^*). \tag{6.6}$$

To prove this, we assume that $J(N; x_0, \lambda)$ attains the minimum at $\lambda = \lambda^*$. Then there exists another admissible sequence $\bar{u} = \{\bar{u}_t\}$ satisfying

$$\lambda_t^* f_t(x_t(\lambda^*), u_t) + (1 - \lambda_t^*) f_t(x_t(\lambda^*), u_t^*) = f_t(x_t(\lambda^*), \bar{u}_t), \tag{6.7}$$

$$\lambda_t^* \alpha_t(x_t(\lambda^*), u_t) + (1 - \lambda_t^*) \alpha_t(x_t(\lambda^*), u_t^*) \geqslant \alpha_t(x_t(\lambda^*), \bar{u}_t). \tag{6.8}$$

Noting that $x_t(\lambda^*) = x_t(x_0, \lambda^*) = x_t(x_0, \bar{u})$ and taking into account of (6.8), we find

$$J(N; x_0, \lambda^*) \geqslant J(N; x_0, \bar{u}). \tag{6.9}$$

Consequently,

$$J(N; x_0, \lambda^*) = J(N; x_0, u^*). \tag{6.10}$$

Now we take λ^* such that $\lambda_t^* = 0$ for $t = 0, 1, \ldots, N-1$ and apply Corollary to the above mentioned problem. Then we have

$$\frac{\partial[\lambda_t^* H_t(x_t^*, p_t^*, u_t) + (1 - \lambda_t^*) H_t(x_t^*, p_t^*, u_t^*)]}{\partial \lambda_t^*} (\lambda_t - \lambda_t^*) \leqslant 0.$$

This implies (6.2).

7. Counter-Example

Consider the problem of minimizing $J(2; x_0, u)$ subject to

$$x_{t+1} = x_t + u_t ,$$

$$J(2; x_0, u) = \sum_{t=0}^{1} [2x_t^2 - u_t^2],$$

$$x_0 = 0, \qquad |u_t| \leqslant 1.$$

By the easy calculation, the optimal decisions are

$$u_0^* = 0, \qquad u_1^* = 1 \quad \text{or} \quad -1.$$

Along these decisions, the Hamiltonian becomes

$$H_0(x_0^*, p_0^*, u_0) = u_0^2.$$

This implies that H_0 does not attain the local maximum at the optimal decision $u_0^* = 0$.

Acknowledgment

The author wishes to express his sincere appreciation to Professor Jin-ichi Nagumo of the University of Tokyo for his helpful suggestions, direction and encouragement.

References

1. S. S. L. Chang. Digitized maximum principle. *Proc. IRE* **48** (1960), 2030-2031.
2. S. Katz. Discrete version of Pontryagin's maximum principle. *J. Electronics Control.* **13** (1962), 179-184.
3. A. G. Butkovskii. The necessary and sufficient conditions for optimality of discrete control systems. *Automation Control.* **24** (1964), 963-970.
4. H. Halkin. Optimal control for systems described by difference equations. *In* "Advances in Control Systems: Theory and Applications," Vol. 1, pp. 173-196. Academic Press, N.Y., 1964.
5. B. W. Jordan and E. Polak. Theory of a class of discrete control systems. *J. Electronics Control.* **17** (1964), 697-711.
6. L. D. Berkovitz. Variational methods in problems of control and programming. *J. Math. Anal. Appl.* **5** (1961), 145-169.
7. R. Bellman. "Dynamic Programming." Princeton University Press, Princeton, N.J., 1957.
8. A. F. Fillipov. On certain questions in the theory of optimal control. *J. SIAM, Control, Ser. A.*, **1** (1962), 76-84.
9. S. Arimoto. On the optimization problem for sampled-data control systems. *Proc. 13th Japan Natl. Congr. Appl. Mech.*, pp. 244-247, University of Tokyo, Tokyo, Japan, 1963.